where the root set t_n differs from the polynomial (18) root set $\tilde{z}_n$ in the p root transpositions $\tilde{z}_k \to 1/\tilde{z}_k^* = t_k$. By deforming the integration contour in (28) one may express all the $\exp(-W_m + i\varphi_m)$ as the sums of $(p+1)$ deductions in the integrand poles, the deduction at infinity being equal to zero, and obtain the result

$$\exp(-W_m + i\varphi_m)$$

$$= \exp[-W + i\alpha(m - N - 1)]$$

$$\times \prod_k^p \frac{(1-\tilde{z}_k)[1 - z_k^* \exp(-i\alpha)]}{(1-\tilde{z}_k^*)[\exp(-i\alpha) - \tilde{z}_k]}$$

$$\times \left[1 - \sum_k^p \frac{\exp(-im\alpha)}{\tilde{z}_k^m} \frac{|\tilde{z}_k|^2 - 1}{\tilde{z}_k^* \exp(-i\alpha) - 1}\right.$$

$$\left. \times \prod_{l \neq k}^p \frac{\exp(-i\alpha) - \tilde{z}_l}{\exp(-i\alpha) - 1/\tilde{z}_l^*} \frac{\tilde{z}_k - 1/\tilde{z}_l^*}{\tilde{z}_k - \tilde{z}_l}\right],$$

where all the sums and products are taken over all the transpositions executed.

As all the bicrystal root magnitudes $|\tilde{z}_k|$ tend to 1, the following expression for Debye-Waller factors may be obtained to first order in $1 - |\tilde{z}_k|$:

$$\exp(-W_m) = \exp(-W)\left\{1 + 2\,\mathrm{Re}\sum_k^p \left[\frac{\exp(-im\alpha)}{\tilde{z}_k^m} \times \frac{1 - |\tilde{z}_k|}{\tilde{z}_k^* \exp(-i\alpha) - 1}\right]\right\}.$$

References

AFANAS'EV, A. M., ALEKSANDROV, P. A., FANCHENKO, S. S., CHAPLANOV, V. A. & YAKIMOV, S. S. (1986). *Acta Cryst.* A**42**, 116-122.

AFANAS'EV, A. M., ALEKSANDROV, P. A. & IMAMOV, R. M. (1986). *Rentgenovskay Structurnay Diagnostica v Issledovanii Pripoverchnostnich Sloyev Monocrystallov.* Moscow: Nauka.

AFANAS'EV, A. M., ALEKSANDROV, P. A., IMAMOV, R. M., LOMOV, A. A. & ZAV'ALOVA, A. A. (1984). *Acta Cryst.* A**40**, 352-355.

AFANAS'EV, A. M. & FANCHENKO, S. S. (1986). *Dokl. Akad. Nauk SSSR*, **287**, 1395-1399.

AFANAS'EV, A. M., KOVALCHUCK, M. V., KOVEV, E. K. & KOHN, V. F. (1977). *Phys. Status Solidi A*, **42**, 415-420.

AFANAS'EV, A. M., KOVALCHUCK, M. V., LOBANOVICH, E. F., IMAMOV, R. M., ALEKSANDROV, P. A. & MELKONYAN, M. K. (1981). *Sov. Phys. Crystallogr.* **26**, 13-20.

AWANO, H., SPERIOSU, V. S. & WILTS, C. H. (1984). *J. Appl. Phys.* **55**, 3043-3048.

BONSE, V., HART, M. & SCHWUTTKE, G. H. (1969). *Phys. Status. Solidi*, **33**, 361-366.

BURGEAT, J. & COLELLA, R. (1969). *J. Appl. Phys.* **40**, 3505-3509.

BURGEAT, J. & TAUPIN, D. (1967). *Acta Cryst.* A**24**, 99-103.

EISENBERGER, A. M., ALEXANDROPOULOS, N. G. & PLATZMAN, P. M. (1972). *Phys. Rev. Lett.* **28**, 1519-1525.

HALLIWELL, M. A. G., LYONS, M. H. & HILL, M. J. (1984). *J. Cryst. Growth*, **68**, 523-531.

KAMENOU, K., HIRAI, I., ASAMA, K. & SAKAI, M. (1979). *J. Appl. Phys.* **34**, 539-542.

QUILLEC, M., GOLDSTEIN, L., LE ROUX, G., BURGEAT, J. & PRIMOT, J. (1984). *J. Appl. Phys.* **55**, 2904-2909.

SPERIOSU, V. S., GLASS, H. L. & KOBAYASHI, T. (1979). *Appl. Phys. Lett.* **34**, 539-542.

TAKAGI, S. (1962). *Acta Cryst.* **15**, 1311-1316.

TAUPIN, D. (1964). *Bull. Soc. Fr. Minéral. Crystallogr.* **87**, 469-475.

YAKIMOV, S. S., CHAPLANOV, V. A., AFANAS'EV, A. M., ALEKSANDROV, P. A., IMAMOV, R. M. & LOMOV, A. A. (1984). *Pis'ma Zh. Eksp. Teor. Fiz.* **39**, 3-5.

YOGI, K., MIYAMOTO, N. & NISHIZAVA, J. (1970). *J. Appl. Phys.* **9**, 246-250.

ZAV'ALOVA, A. A., IMAMOV, R. M., LOMOV, A. A., MARGUSCHEV, Z. C. & MASLOV, A. V. (1987). *Krystallografiya.* In the press.

Acta Cryst. (1988). A**44**, 33-37

The Statistical Significance of Difference Densities

BY E. N. MASLEN

Crystallography Centre, University of Western Australia, Nedlands 6009, *Western Australia*

(*Received* 31 *March* 1987; *accepted* 11 *August* 1987)

Abstract

The statistical properties of a difference density $\Delta\rho$ are not fully characterized by the standard deviation $\sigma(\Delta\rho)$, which relates to the density at a point. That is not sufficient information to assess the significance accurately for the density within a finite volume. The reliability of a complete $\Delta\rho$ map may be determined by applying standard statistical tests to the chi-square index

$$\chi^2 = \sum_{\mathbf{s}} \sigma^{-2}(\mathbf{s})[\Delta F(\mathbf{s})]^2$$

from a least-squares refinement, where ΔF is a structure-factor residual and σ^2 is the variance in the structure factor, or equivalently to the goodness-of-fit

0108-7673/88/010033-05$03.00

index $[(\chi^2/\nu-1)/2]^{1/2}$, where ν is the number of degrees of freedom in the structure refinement. A similar treatment applies to component volumes or features in the difference density for which the chi-square index is

$$\chi^2=\sum_{\mathbf{s}}\sigma_n^{-2}(\mathbf{s})[\Delta F_n(\mathbf{s})]^2,$$

where ΔF_n is obtained by Fourier transformation from the nth component of $\Delta\rho$ and σ_n^2 is the variance rescaled by the fraction of the cell volume occupied by the feature.

Significance and $\sigma^2(\Delta\rho)$

In scattering experiments we probe the structure of matter. The measurements are of limited value without reliable estimates of their precision. This is equally true for discrete quantities derived from the measurements, such as interatomic distances, which can be compared using significance criteria derived by standard methods. It also applies to the images of scattering density $\rho(\mathbf{r})$ obtained by inverse Fourier transformation from the phased scattering amplitudes. For scattering from a crystal that transform is the summation

$$\rho(\mathbf{r})=V^{-1}\sum_{\mathbf{s}}\exp(-2\pi i\mathbf{s}\,.\,\mathbf{r})F(\mathbf{s}) \qquad (1)$$

where V is the unit-cell volume and $F(\mathbf{s})$ is the structure factor.

It is more convenient to study the difference density, ideally obtained by subtracting from the exact $\rho(\mathbf{r})$ the density $\rho_c(\mathbf{r})$ for a reference model resembling the true structure, but in practice obtained by evaluating

$$\Delta\rho(\mathbf{r})=V^{-1}\sum_{\mathbf{s}}\exp(-2\pi i\mathbf{s}\,.\,\mathbf{r})\Delta F(\mathbf{s})$$

where

$$\Delta F(\mathbf{s})=F_o(\mathbf{s})-F_c(\mathbf{s}). \qquad (2)$$

$F_o(\mathbf{s})$ is the measured structure factor, and the summation ranges over the finite set for these structure factors. $F_c(\mathbf{s})$ is the corresponding structure factor for the reference model.

We confine our attention to centrosymmetric structures and assume that in a scattering experiment the phases of the structure factors are known exactly. The errors in the measured structure factors $\delta F(\mathbf{s})$ are assumed to be normally distributed about zero with variance $\sigma^2(\mathbf{s})$. The reference model for $\rho_c(\mathbf{r})$ is also treated as exact and the difference density is analysed on that basis. That is, if all the $\delta F(\mathbf{s})$ were zero the difference density would be a representation, correct at that resolution, of the difference between the true scattering density and that of the particular reference model. The covariance terms for the structure factors are neglected.

When analysing scattering experiments it is often desirable to consider $\Delta\rho$ as containing a set of components $\{\Delta\rho_n\}_{n=1,N}$ such that

$$\Delta\rho=\sum_{n=1}^{N}\Delta\rho_n. \qquad (3)$$

A component might be due to an error in a nuclear position, for example, or to the neglect of anharmonicity in the thermal motion of a particular atom. The objective when analysing difference densities is to differentiate such components from the effects of errors in the measured structure factors. Significance tests can assist that analysis by identifying parts of the difference density which cannot reasonably be attributed to noise in the structure factors. Significance here has its standard meaning in statistical inference, namely the result of testing whether the quantity (the difference density) could have arisen by chance from the random errors in the measurements [*i.e.* the $\delta F(\mathbf{s})$].

The variance at a point in the difference density can be derived as in equation (17) of Rees (1977) which reduces to

$$\sigma^2(\Delta\rho)=\sum_i\left[2V^{-1}\sum_e\cos(2\pi\mathbf{s}_{ie}\,.\,\mathbf{r})\right]^2\sigma^2(\mathbf{s}) \qquad (4)$$

where the first sum is over the independent reflections and the second is over the symmetry equivalents in one hemisphere. Unless $\mathbf{r}$ is near a special position, or the number of structure factors is small, the term in brackets may be approximated by replacing the square of each cosine function by its effective mean value of $\frac{1}{2}$ without serious loss of accuracy.

The maximum $|\Delta\rho|$ within a component is often regarded as indicating its level of significance. The maximum of the ratio $|\Delta\rho|/\sigma(\Delta\rho)$ is treated as if it were equivalent to the ratio $|\Delta x|/\sigma(x)$ for the one-dimensional random variable x, which is the square root of minus twice the argument for the exponential term in a Gaussian probability function. However it is not obvious how to determine the significance of components accurately from $|\Delta\rho|/\sigma(\Delta\rho)$ because it is an estimate based on one point only.

Accurate assessment of the significance of a finite volume of difference density $\Delta\rho$ in real space is especially difficult because of the mathematical complexity arising from the covariance and higher-order correlation terms. Because of the equivalence between real- and reciprocal-space representations, however, the statistical significance of a probability function is not altered by a Fourier transformation (Cramér, 1946). The significance of the complete $\Delta\rho$ measurement in real space is identical to that of the corresponding set of structure-factor differences $\{\Delta F(\mathbf{s})\}$ in reciprocal space. The latter is much simpler to evaluate because the set of structure factors is discrete. In practical applications the set of structure factors

is also finite. Furthermore, when the structure factors are determined from independently measured intensities, the covariance and higher-order correlation terms are negligible for most purposes.

The significance of the set $\{\Delta F(\mathbf{s})\}$ can be related to chi-square indices, which are conveniently calculated during least-squares refinement of a crystal structure. The sample estimate of the chi-square index is

$$\chi^2 = \sum_{\mathbf{s}} \sigma^{-2}(\mathbf{s})[\Delta F(\mathbf{s})]^2. \quad (5)$$

The probability function for χ^2 (Cramér, 1946) is

$$P(\chi^2, \nu) = (\chi^2/2)^{\nu/2-1} \exp(-\chi^2/2)/2\Gamma(\nu/2), \quad (6)$$

where ν, the number of degrees of freedom, is the number of independent reflections minus the number of variable parameters in a least-squares refinement.

The probability that the chi-square index exceeds a given value by chance is listed as a function of ν by Cramér (1946) in his Table 3. The probability that the reduced chi-square index, χ^2/ν, exceeds a given value by chance is listed in his Table C-4 by Bevington (1969).

The chi-square index with ν degrees of freedom is asymptotically normal, with a mean value ν and variance 2ν, for ν large (Cramér, 1946), in which case (6) becomes

$$P(\chi^2, \nu) = [1/(4\pi\nu)^{1/2}] \exp[-(\chi^2 - \nu)/4\nu]. \quad (6a)$$

The probability that χ^2 exceeds ν by more than $\lambda(2\nu)^{1/2}$ by chance is listed in his Table 2 by Cramér (1946). The number of degrees of freedom in most crystallographic experiments is large enough for the asymptotic form (6*a*) to apply. On the other hand, the derivation of the probability functions [(6) and (6*a*)] assumes that the residuals are linear functions of the parameters, which only holds approximately in practice. Within the limits of that approximation the significance of difference densities can be assessed by applying standard methods of inference to normal distributions.

Significance of components

We now extend the reasoning of the previous section to the components $\Delta\rho_n$. For simplicity we first consider the case where the components divide the cell into a finite set of non-overlapping fragments $\{V_n\}_{n=1,N}$. Corresponding to (3) in real space we have in reciprocal space

$$\Delta F = \sum_{n=1}^{N} \Delta F_n \quad (7)$$

where ΔF_n is related to $\Delta\rho_n$ by the Fourier transformation

$$\begin{aligned}\Delta F_n &= \int_{\mathbf{V}} \Delta\rho_n \exp(2\pi i \mathbf{s}\,.\,\mathbf{r})\, d\tau \\ &= \int_{\mathbf{V}_n} \Delta\rho_n \exp(2\pi i \mathbf{s}\,.\,\mathbf{r})\, d\tau \\ &= \int_{\mathbf{V}_n} \Delta\rho \exp(2\pi i \mathbf{s}\,.\,\mathbf{r})\, d\tau. \end{aligned} \quad (8)$$

$\mathbf{V}$ denotes integration over the full cell. The alternative forms are valid because $\Delta\rho_n$ vanishes outside $\mathbf{V}_n$.

The significance of the set of component structure-factor differences $\{\Delta F_n(\mathbf{s})\}_{n=1,N}$ is given by standard statistical tests applied to the chi-square indices for the component

$$\chi_n^2 = \sum_{\mathbf{s}} \sigma_n^{-2}(\mathbf{s})(\Delta F_n)^2 \quad (9)$$

where ΔF_n is defined in (8) and σ_n^2 is the dispersion of the probability function for the error in ΔF_n. Since ΔF_n can be calculated from $\Delta\rho_n$ by a numerical Fourier transformation, the chi-square index can be evaluated provided $\sigma_n^2(\mathbf{s})$ can be determined.

The errors in the structure factors $\{\delta F(\mathbf{s})\}$ may be expressed in terms of a real-space error function $\delta\rho(\mathbf{r})$ such that

$$\begin{aligned}\delta F(\mathbf{s}) &= \int_{\mathbf{V}} \delta\rho(\mathbf{r}) \exp(2\pi i \mathbf{s}\,.\,\mathbf{r})\, d\tau \\ &= \sum_{n=1}^{N} \int_{\mathbf{V}_n} \delta\rho(\mathbf{r}) \exp(2\pi i \mathbf{s}\,.\,\mathbf{r})\, d\tau \\ &= \sum_{n=1}^{N} \delta F_n(\mathbf{s}), \end{aligned} \quad (10)$$

which defines $\delta F_n(\mathbf{s})$. The variance $\sigma_n^2(\mathbf{s})$ is the dispersion of the probability function for $\delta F_n(\mathbf{s})$.

As the structure is centrosymmetric, elements of $\mathbf{V}_n$ in (10) can be combined so that the geometrical term $\exp(2\pi i \mathbf{s}\,.\,\mathbf{r})$ reduces to a cosine function. The dispersion of the function $\delta\rho(\mathbf{r})$ is uniform through the cell except for modest increases (up to a factor of two) near a special position, and its sign fluctuates randomly over distances greater than the resolution. Similar comments apply to the cosine multiplier. A preliminary hypothesis is that the integrals in (10) can be recast as finite sums satisfying the central limit theorem (Cramér 1946), and thus the $\delta F_n(\mathbf{s})$ have a normal distribution.

We denote the volume of the fragment $\mathbf{V}_n$ by V_n. It is helpful to begin with the case where V/V_n is independent of n, *i.e.* V_n is V/N for all components. If our preliminary hypothesis is satisfied the δF_n combine to form δF as do the steps in a random walk (Reif, 1965), in which case

$$\sigma_n^2(\mathbf{s}) = N^{-1}\sigma^2(\mathbf{s}). \quad (11)$$

If one relaxes the restriction on V_n to $V_n < V$, (11) becomes

$$\sigma_n^2(\mathbf{s}) = (V_n/V)\sigma^2(\mathbf{s}) \quad (12)$$

so that the chi-square index for the nth component is

$$\chi_n^2 = (V/V_n) \sum_{\mathbf{s}} \sigma^{-2}(\mathbf{s})(\Delta F_n)^2, \qquad (13)$$

from which the significance of the nth component may be calculated by standard methods.

The derivation of (11) is suspect if the component volumes are small compared with the resolution, in which case the δF_n from neighbouring volumes may be positively correlated, rather than randomly related. Similar difficulties arise if subdivision of the difference density is extended to overlapping fragments, with $\Delta\rho$ partitioned using non-negative weighting factors in the overlapping region. The contributions of overlapping fragments will again be positively correlated.

The consequences are indicated by examining the extreme case in which all the δF_n are identical and so add in phase. In (11) the multiplier N^{-1} is replaced by N^{-2} and the multiplier (V/V_n) in (13) becomes $(V/V_n)^2$. Thus if the ΔF_n are positively correlated the true chi-square is greater than that given by (13). The use of (13) as an index thus sets *a lower limit* to the level of significance. The deficiencies of this method are fail safe, and it is unlikely that they will be serious in practical situations.

Nevertheless, a word of caution may be appropriate. This argument breaks down if $\Delta\rho$ is partitioned with negative weights so that a volume element makes a contribution which is negative to one component, and positive to another. Such use of negative weights in partitioning $\Delta\rho$ could result in negative correlation. The chi-square indices predicted by (13) for negatively correlated components would be too large.

Example

The properties of the foregoing theory may be illustrated by applying it to a test case, namely silicon. The difference densities shown in Fig. 1 were derived from experimental structure factors measured by Teworte & Bonse (1984), supplemented by the 222 reflection of Alkire, Yelon & Schneider (1982). The data, which are on an absolute scale, include all ten independent reflections with $|\mathbf{s}| = (2\sin\theta)/\lambda < 1{\cdot}06$ Å^{-1}, and seven more structure factors with $1{\cdot}06 < |\mathbf{s}| < 2{\cdot}10$ Å^{-1}.

The calculated structure factors were based on the isolated atom model evaluated using atomic scattering factors, derived from the wave functions of Clementi (1965), and the dispersion measurements of Deutsch & Hart (1985), and were corrected for Thomson scattering from the nucleus. The temperature factor coefficient B was fixed at 0·464 Å^2, as determined from an earlier multipole refinement.

The difference density will contain, among other real contributions, a component caused by an inaccuracy in that value. The sections of the $\Delta\rho$ and $\sigma(\Delta\rho)$ maps plotted in Fig. 1, which contain the mean positions for a pair of bonded silicon atoms, were evaluated using the full data set (*a* and *b*) and the ten reflections with $|\mathbf{s}| < 1{\cdot}06$ Å^{-1} (*c* and *d*). The limited range of contours in map (*b*) shows that $\sigma^2(\Delta\rho)$ is

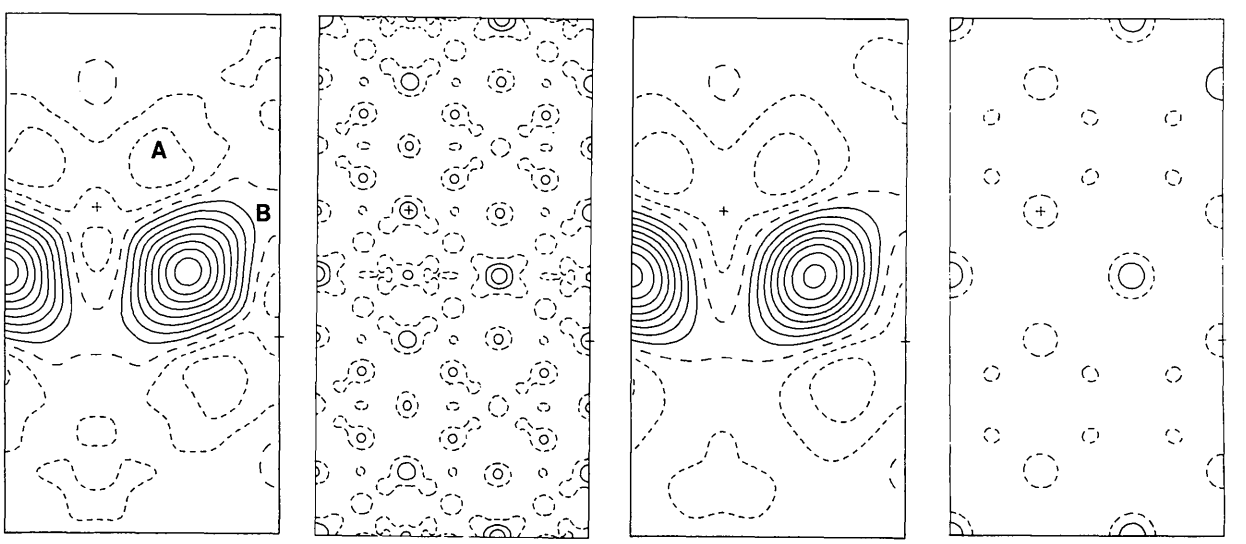

Fig. 1. Sections of density maps for silicon. (*a*) $\Delta\rho$ for 17 reflections with $|\mathbf{s}| < 2{\cdot}10$ Å^{-1}. (*b*) $\sigma(\Delta\rho)$ for map (*a*) with contours ranging from 0·004 (dashed) to 0·005 e Å^{-3}. (*c*) $\Delta\rho$ for ten reflections with $|\mathbf{s}| < 1{\cdot}06$ Å^{-1}. (*d*) $\sigma(\Delta\rho)$ for map (*c*) with contours ranging from 0·003 (dashed) to 0·004 e Å^{-3}. The mean positions for bonded silicon atoms are shown as crosses. The features discussed in the text are labelled *A* and *B* in (*a*).

Table 1. *Probability data from difference densities for silicon*

Term	Density component	10 reflections ($\nu = 9$)	17 reflections ($\nu = 16$)
χ^2	Whole map	35 736	35 767
	A	23 534	23 672
	B	114 020	114 170
$[(\chi^2/\nu - 1)/2]^{1/2}$	Whole map	44·5	33·4
	A	36·2	27·2
	B	79·6	59·7
$\lvert\Delta\rho\rvert/\sigma(\Delta\rho)$	*A*	26·2	17·9
	B	54·6	39·7

approximately uniform even though there are only 17 independent structure factors. Since the *B* value is the only free parameter in the model there are 16 and nine degrees of freedom for the full and low-angle data sets respectively.

Chi-square indices were evaluated from the silicon deformation densities plotted in Fig. 1 for (i) the whole map, (ii) the negative feature, $-0{\cdot}06$ e Å^{-3} in depth, 0·5 Å from the Si nucleus on the extrapolation of the Si–Si bond, labelled *A* in Fig. 1(*a*), and (iii) a positive feature 0·21 e Å^{-3} high centred on the Si–Si bond and labelled *B* in Fig. 1(*a*). Boundaries were chosen to approximate the contours at $-0{\cdot}025$ and $+0{\cdot}025$ e Å^{-3} for features *A* and *B* respectively. The regions within the unit cell which are equivalent by symmetry were included when calculating ΔF_n and V_n. In one unit cell there are 32 slightly overlapping regions each with volume 0·82 Å^3 for the negative feature (*A*) and 16 repetitions, each with volume 1·63 Å^3 for the positive feature (*B*). The differences in boundaries and probabilities on changing from the truncated to the full data set were slight. The chi-square indices are listed in Table 1.

The goodness-of-fit index $[(\chi^2/\nu - 1)/2]^{1/2}$, which is the square root of minus twice the argument of the exponential factor in the probability function (6*a*), is the analogue of the maximum $|\Delta\rho|/\sigma(\Delta\rho)$ used as a point estimate of significance. Both values are included in Table 1. $[(\chi^2/\nu - 1)/2]^{1/2}$ is larger, as expected, since the probability term is obtained by integration over a volume which is large compared to the resolution, whereas the maximum of $|\Delta\rho|/\sigma(\Delta\rho)$ relates to a single point.

The results in Table 1 emphasize another aspect of significance tests applied to difference densities. The maximum value of $|\Delta\rho|/\sigma(\Delta\rho)$ within feature *A* decreases from 26·2 in the low-angle map (*c*) to 17·9 in the full-angle map (*a*). There are analogous decreases in magnitude of the arguments for the other probability functions. It is axiomatic that adding information which is not inconsistent with an initial set of data cannot reduce its significance. The probability functions corresponding to the arguments in the third and fourth columns of Table 1 are not different approximations to the same quantity. The significance levels derived from them apply to the respective 10 to 17 reflection images explicitly. The 10-reflection map contains almost all the useful information. The significance levels for the 17-reflection map relate, not just to the general appearance of the features, but also to how well the fine structure associated with the high-angle reflections is described.

In so far as valence scattering is a low-Bragg-angle phenomenon, the significance tests from extensive data sets may yield inaccurate estimates of the reliability of features associated with the valence density. This applies particularly to estimates based on the maximum $|\Delta\rho|/\sigma(\Delta\rho)$ ratio. For sharp features dominated by high-angle data, such as those due to thermal anharmonicity, the significance estimates based on that ratio may be useful. Where a component has dimensions much larger than the resolution, however, the single-point estimate will be inaccurate, in the same way that the 17-reflection value of $|\Delta\rho|/\sigma(\Delta\rho)$ is a poor approximation to the 10-reflection $[(\chi^2/\nu - 1)/2]^{1/2}$ value.

This work was supported by the Australian Research Grants Scheme.

References

ALKIRE, R. W., YELON, W. B. & SCHNEIDER, J. R. (1982). *Phys. Rev. B*, **26**, 3097–3104.

BEVINGTON, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill.

CLEMENTI, E. (1965). *IBM J. Res. Dev.* **9**, Supplement 2.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

DEUTSCH, M. & HART, M. (1985). *Acta Cryst.* A**41**, 48–55.

REES, B. (1977). *Isr. J. Chem.* **16**, 180–186.

REIF, F. (1965). *Fundamentals of Statistical and Thermal Physics*, pp. 25–40. New York: McGraw-Hill.

TEWORTE, R. & BONSE, U. (1984). *Phys. Rev. B*, **29**, 2102–2108.

Acta Cryst. (1988). A**44**, 38-45

Application of the Molecular Replacement Method to Multidomain Proteins. 1. Determination of the Orientation of an Immunoglobulin Fab Fragment

BY M. CYGLER AND W. F. ANDERSON

MRC Group on Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

(*Received* 15 *April* 1987; *accepted* 11 *August* 1987)

Abstract

Multidomain proteins provide special problems in the application of the molecular replacement method of structure determination. The structure of the Fab fragment from the autoimmune poly(dT)-specific antibody HED10 has been determined using molecular replacement. An analysis of the effects of varying the model and the parameters used in the rotation function indicates that dividing the molecule into individual relatively rigid domains simplifies interpretation of the results, and that the optimal parameters depend on the molecule under study.

Introduction

The continually growing number of proteins whose three-dimensional structures have been determined increases the possibility that a new protein being investigated will have some features in common with one or more of the known structures. In addition, there is growing interest among biochemists and protein crystallographers in determining the changes in protein folding and/or packing caused by specific modifications of the amino-acid sequence. For these cases the application of the standard multiple isomorphous replacement technique (Blundell & Johnson, 1976) to determine phases, while it will give the final answer, may not be the fastest or most straightforward way to achieve this goal. A more direct approach is to utilize information from a closely related protein of known structure by application of the molecular replacement (MR) technique (Rossmann, 1972). If successful, this approach can decrease dramatically the time required to determine the protein structure and can make heavy-atom derivatives unnecessary. While the application of the MR method is computationally intensive, this is no longer an obstacle.

The task of positioning a model molecule in the unit cell involves six degrees of freedom: three to determine the orientation and three to determine the translation of the molecule. From a theoretical analysis of the properties of the Patterson function (Hoppe, 1957; Rossmann & Blow, 1962) it became obvious that such a task can be reduced to two consecutive three-dimensional problems. The first step is the determination of the correct orientation of the model and the second is the determination of the position of the correctly oriented model within the unit cell. The orientation of the molecule is determined by the comparison of the Patterson function of the unknown crystal with that of the model molecule in all possible orientations. The most widely used score function for analyzing the similarity of Patterson functions is the integral of their products over a volume around the origin of the unit cell. Fast algorithms that can be used to calculate this rotation function (RF) in both direct (Huber, 1965; Steigemann, 1974) and reciprocal (Crowther, 1972) space have been developed and have been successfully applied.

Despite many years of experimenting with the RF in a number of laboratories there is no clear understanding of the effect of various factors involved in the calculations on the final success or failure of the method. The current approach is to repeat the calculations many times, varying parameters that are considered important by the investigator.

There are numerous examples of successful applications of the MR method [*e.g.* lysozyme (Bott & Sarma, 1976); insulin (Dodson, Harding, Hodgkin & Rossmann, 1966); hemoglobin (Derewenda, Dodson, Dodson & Brzozowski, 1981); serine proteases (Fujinaga, Read, Sielecki, Ardelt, Laskowski & James, 1982; McPhalen, Svendsen, Jonassen & James, 1985); phospholipase A2 (Dijkstra, van Nes, Kalk, Brandenburg, Hol & Drenth, 1982); immunoglobulin pFc′ fragment (Phizackerley, Wishner, Bryant, Amzel, Lopez de Castro & Poljak, 1979); phycocyanin (Schirmer, Huber, Schneider, Bode, Miller & Hackert, 1986)] and some methodological and practical aspects have been the subject of a special symposium (Daresbury Study Weekend, 1985). The procedure is relatively straightforward in the case of a rigid molecule and success depends primarily on how well the model approximates the unknown protein.

Multidomain or multisubunit proteins that undergo a conformational change upon ligand binding [*e.g.* hemoglobin, hexokinase, arabinose binding protein, citrate synthase, *etc.* (Huber & Bennett, 1983)] or that are relatively flexible [*e.g.* immunoglobulins (Amzel

0108-7673/88/010038-08$03.00